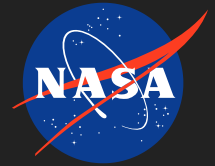# Mining and Utilizing Dataset Relevancy from Oceanographic Dataset Metadata, Usage Metrics, and User Feedback to Improve Data Discovery and Access (MUDROD)

Completed Technology Project (2015 - 2017)

Project Introduction

We propose to mine and utilize the combination of Earth Science dataset, metadata with usage metrics and user feedback to objectively extract relevance for improved data discovery and access across a NASA Distributed Active Archive Center (DAAC) and other data centers. As a point of reference, the Physical Oceanographic Distributed Active Archive Center (PO.DAAC) aims to provide datasets to facilitate scientists in selecting Earth observation data that fit better their needs in various aspects of Physical Oceanography. The TRL 5 technology of data relevance mining, developed by George Mason University (GMU), NASA, and U.S. Geological Survey (USGS) to support the Geosearch operation and contributed as open source through GeoNetwork, will be improved and tested within the PO.DAAC's metadata-centric discovery system as a TRL 7 technology upon completion of the project. This project will focus on the following objectives and activities: ' Integrating and interfacing the data relevance mining and utilizing the data relevance mining system to include the functionality of a) dataset relevance reasoning based on Jena, an open source semantic reasoning engine, b) dataset similarity calculation, c) recommendations based on dataset metadata attributes and user workflow patterns, and d) ranking results based on similarity between user search terms and dataset usage contexts. ' Leveraging the PO.DAAC data science expertise and user communities to a) capture the ocean science data context and record relevant dataset relevance metrics as triple stores, b) analyze and mine user search and download patterns, c) test the developed system in an experimental environment, d) integrate the system into the PO.DAAC testbed and test the feasibility of integration for open usage and feedback. ' Laying the groundwork for an objective mining and extraction service for data relevance with other data search and discovery systems, such as ECHO, GEOSS clearinghouse, and Data.gov, for data sharing across NASA and non-NASA data systems. The proposed technology has the potential to enhance the NASA Earth Science data discovery experience by more efficiently and objectively providing scientists with the ability to discover and select the datasets most relevant to their scope of interest. TRL: Existing Data Relevance Mining and Usage/TRL5. We expect a TRL 7 exit of the technology through the two year research Keywords: PO.DAAC, Data Relevance, Mining, Reasoning, Ranking, Recommendation
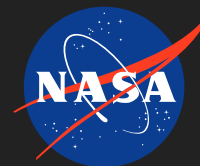

Mining and Utilizing Dataset Relevancy from Oceanographic Dataset Metadata, Usage Metrics, and User Feedback to Improve Data Discovery and Access
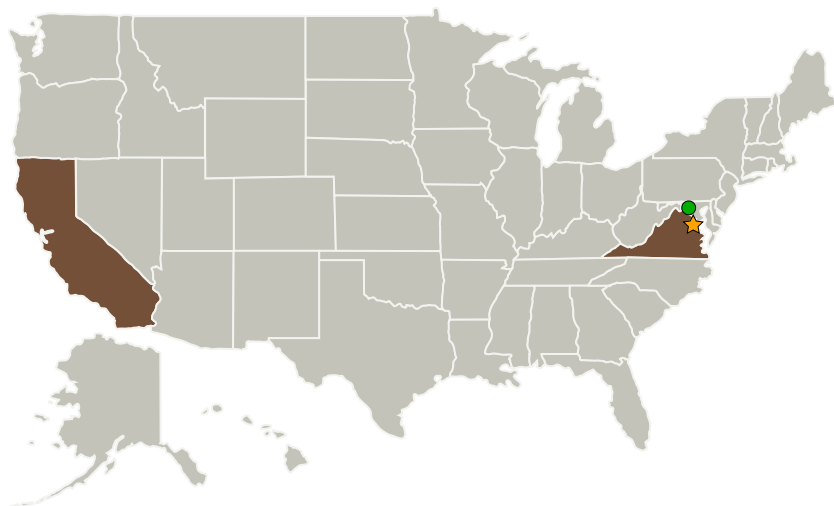
Table of Contents

Mining and Utilizing Dataset Relevancy from Oceanographic Dataset Metadata, Usage Metrics, and User Feedback to Improve Data Discovery and Access (MUDROD)
Completed Technology Project (2015 - 2017)

Primary U.S. Work Locations and Key Partners

Organizations Performing Work	Role	Type	Location
⭐NASA Headquarters(HQ)	Lead Organization	NASA Center	Washington, District of Columbia
George Mason University	Supporting Organization	Academia	Fairfax, Virginia
🟢Goddard Space Flight Center(GSFC)	Supporting Organization	NASA Center	Greenbelt, Maryland

Primary U.S. Work Locations	
California	Virginia

Organizational Responsibility

Responsible Mission Directorate:

Science Mission Directorate (SMD)

Lead Center / Facility:

NASA Headquarters (HQ)

Responsible Program:

Advanced Information Systems Technology

Project Management

Program Director:

Pamela S Millar

Program Manager:

Jacqueline J Le Moigne

Principal Investigator:

Chaowei Yang

Co-Investigator:

Carol-ann Courtney

TechPort

Printed on 11/29/2022
09:13 PM UTC

For more information and an accessible alternative, please visit:
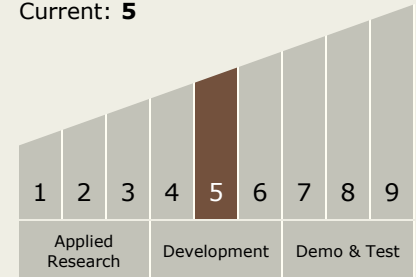https://techport.nasa.gov/view/18711

Page 2

Mining and Utilizing Dataset Relevancy from Oceanographic Dataset Metadata, Usage Metrics, and User Feedback to Improve Data Discovery and Access (MUDROD)
Completed Technology Project (2015 - 2017)

Technology Maturity (TRL)

Start: **5**
Current: **5**

1	2	3	4	5	6	7	8	9
Applied Research			Development			Demo & Test		

Technology Areas

Primary:

- TX11 Software, Modeling, Simulation, and Information Processing
  └ TX11.4 Information Processing
    └ TX11.4.2 Intelligent Data Understanding

Target Destination
Earth